Pseudorandomness, symmetry, smoothing: II (draft)

Harm Derksen*
Northeastern University
ha.derksen@northeastern.edu

Peter Ivanov†
Northeastern University
ivanov.p@northeastern.edu

Chin Ho Lee
North Carolina State University
chinho.lee@ncsu.edu

Emanuele Viola‡
Northeastern University
viola@ccs.neu.edu

May 20, 2024

1 Introduction and our results

This is a draft of a companion paper to [DILV24], to which we refer for background and discussion.

A distribution D over $\{-1,1\}^n$ is (ϵ,k)-*biased* if for every $S \subseteq [n]$ of size $0 < |S| \leq k$ we have $|\mathbb{E}[D^S]| \leq \epsilon$, where $D^S := \prod_{i\in S} D_i$. If $\epsilon = 0$ then any k bits are uniform and D is called k-*wise uniform* or simply k-*uniform*. If $k = n$ then D is called ϵ-*biased*. The study of these distributions permeates and precedes theoretical computer science. They were studied already in the 40's [RR47], are closely related to universal hash functions [CW79], error-correcting codes (see e.g. [HH23]), and in their modern guise were introduced in the works [ABI86, CGH+85, NN90].

1.1 Our results: Bounded uniformity

Exploited in countless works, one of the most useful properties of such distributions D is that the distribution of their Hamming weight $|1^\top D|$ is approximately binomial. Yet, perhaps surprisingly, available bounds were either loose or only applied to specific settings of parameters.

We obtain new lower bounds on $\mathbb{P}[|1^\top D| \geq t]$, the cumulative density function of the weight distribution of any k-uniform distribution D, generalizing or strengthening a number of previous works. In particular this allows us to bound the distance between $|1^\top D|$ and binomial.

*Partially supported by NSF grant DMS 2147769.
†Supported by NSF grant CCF-2114116.
‡Supported by NSF grant CCF-2114116.

We summarize our results in a table, then discuss them, and finally state corresponding theorems.

t	Upper bound ($\forall D$, error is $\leq$)	Lower bound ($\exists D$ s.t. error is $\geq$)
Regime 1: $t \in [0, c\sqrt{n}]$	$c/\sqrt{k}$ [BGGP12, DGJ+10, DKN10]	$c/\sqrt{k}$ (Theorem 1)
Regime 2: $t \in [c\sqrt{n}, c\sqrt{nk}]$	$c^{t^2/n}$ (Folklore, Corollary 11)	$c^{t^2/n}/\sqrt{k}$ (Theorem 1)
Regime 3: $t \in [c\sqrt{nk}, \infty]$	$(ckn/t^2)^{k/2}$ (Folklore, Corollary 11)	$(ckn/t^2)^{k/2}$ (Theorem 2)

Table 1: Upper and lower bounds on $\max_D |\mathbb{P}[|1^\top D| \geq t] - \mathbb{P}[B \geq t]|$ where D is k-wise uniform on $\{-1,1\}^n$, and B is binomial

Regime 1. We obtain error bounds tight up to constant, in particular removing a logarithmic factor from a lower bound sketched in [BGGP12].

Regime 2. The upper bound follows by applying the tail bound for k'-wise uniformity, for $k' := a \leq k$, which shows that $\mathbb{P}[|1^\top D| \geq t] \leq c^{-a}$. This is also a bound on the error because $\mathbb{P}[B \geq t] \leq c^{-t^2/n}$ as well.

Regime 3. As for Regime 2, we note that the upper bound for the tail is also an upper bound on the error.

The lower bound follows from Theorem 2 because $\mathbb{P}[B \geq t] \leq c^{-t^2/n}$ which less than $1/2$ the bound in Theorem 2 when $t \geq \sqrt{ckn}$. This generalizes [BS15].

To illustrate this regime, note that in particular for $t = c\sqrt{n \log n}$ we show that the error is large: $\geq (c/\log n)^k$, whereas $\mathbb{P}[B \geq t] \leq 1/n^c$ is exponentially smaller.

Theorem 1. *For every k and t, there exists a k-wise uniform distribution D on $\{-1,1\}^n$ such that*

$$\mathbb{P}[1^\top D \geq t] - \mathbb{P}[1^\top U \geq t] \geq c\sqrt{\tfrac{n}{k}}\,\mathbb{P}[1^\top U = t] \geq 2^{-t^2/n}/(2\sqrt{k}).$$

The latter inequality is Fact 7.

Note that $\mathbb{P}[1^\top U = t]$ on the right hand side cannot be replaced with $\mathbb{P}[1^\top U \geq t]$.

Theorem 2. *For every $k \leq (n/9)^{1/3}$ and $t \geq \sqrt{nk}$, there exists a k-wise uniform distribution D on $\{-1,1\}^n$ such that $\mathbb{P}[1^\top D \geq t] \geq \frac{1}{4k^{3/2}}(\frac{kn}{16t^2})^{k/2}$.*

1.2 Our results: Small-bias distributions

We develop a paradigm to obtain small-bias distributions from k-uniform distributions while retaining some of their deviation properties. The paradigm has two steps. First, *symmetrize* the distribution. If the distribution is typically supported on nearly balanced strings, this step has the effect of making the bias small on tests of size not too large. Second, add noise, following [LV17]. This makes the bias small on large tests. Historically, this was our first approach to obtain the main results in [DILV24].

Using this approach we extend the results in the previous subsection to small-bias distributions.

Corollary 3. *For any k and $t \in [c\sqrt{n}, (ckn^3)^{1/4}]$, there exists an $(ck/n)^{k/4}$-biased distribution D such that $\mathbb{P}[1^\top D \geq t] \geq \min\{1/\sqrt{k}, (ckn/t^2)^{k/2}\}$.*

1.3 Our results: Small-bias distributions plus noise

A main result from [DILV24] is that n^{-k}-bias distributions plus noise do not fool thresholds with error less than c^k. This is then used to show that they do not fool, even with constant error, other models like small-space algorithms or small-depth circuits.

In this paper we give four alternative proofs of this main result from [DILV24]. Also, the proofs provide additional information. In [DILV24] the pseudorandom distribution put *more* mass than the binomial on the tail. In some of the proofs presented here, the mass will be *less*. This also provides a more complete picture of how these distributions can be designed.

One proof follows from Corollary 3. The other three proofs are presented in Section 5. These proofs may be of independent interest. One proof uses convex geometry and a new bound between the normal distribution and any mixture of few normal distributions with bounded variance.

2 Proof of Theorem 2

The proof of Theorem 2 follows the same strategy in [BS15]. However, there are some noticeable differences. First, we decouple the threshold and the error parameters. Second, we do not pass the argument to the Gaussian distribution. Finally and most importantly, their proof appears to incur a loss of a $1/\sqrt{n}$ factor in their lower bound on the tail, which is significant in certain regimes of parameters, e.g. $k = 2$ and $t = n^{o(1)}$.

To prove Theorem 2, we use tools in approximation theory. In particular, to remove the loss in [BS15], we rely on the following inequality by Erdéyi.

Lemma 4 (Theorem 2.1 ($q = 1$ case) in [Erd16])**.** *For $m \in \mathbb{N}$ and $L > 0$, let $Q \in \mathbb{C}[x]$ be a degree-d univariate polynomial (with complex coefficients) such that*

$$|Q(0)| > \frac{1}{L}\Big(\sum_{j=1}^{m} |Q(j)|\Big).$$

Then $d \geq 7\sqrt{m/L}$.

We also need the following inequality due to Ehlich, Zeller, Coppersmith, Rivlin, and Cheney.

Lemma 5 (Lemma 20 in [BS15])**.** *Let p be a univariate degree-d polynomial such that $|p(i)| \leq 1$ on $i \in \{0, \ldots, m\}$, where $3d^2 \leq m$. Then $|p(x)| \leq 3/2$ for every $x \in [0, m]$.*

We will also use the following extremal property of Chebyshev polynomials T_k.

Fact 6. *Let p be a univariate polynomial of degree k such that $|p(t)| \leq 1$ on $[-1, 1]$. For every $s \geq 1$, $|p(s)| \leq T_k(s) \leq (2|s|)^k$, where T_k is the Chebyshev polynomial of degree k.*

Proof. Fill in later. □

Fact 7 (cf. Lemma 23 in [BHLV19]). *If a is an integer such that $|a| \le n$ and $a \equiv n \bmod 2$, then $\mathbb{P}[B = a] \ge 2^{-a^2/n} \frac{1}{2\sqrt{n}}$.*

By strong duality, we have

$$\max_D \mathbb{P}\big[\mathbb{1}(1^\top D \ge t)\big] = \min_p \mathbb{E}\big[p(U)\big],$$

where the maximum is over all k-wise uniform distributions D, and the minimum is over all degree-k (upper sandwiching) polynomials $p\colon \{-1,1\}^n \to \mathbb{R}$ such that $p(x) \ge \mathbb{1}(1^\top x \ge t)$.

Let p be a degree-k polynomial attaining $\delta := \min_p \mathbb{E}[p(U)]$. Define the univariate polynomial q to be the symmetrization of p, that is, $q(\sum_{i=1}^n x_i) := p(x)$.

We use Lemma 4 and 5 to bound $q(t)$ on $t \in [-\sqrt{kn}, \sqrt{kn}]$.

Lemma 8. *$|q(t)| \le 2\delta \cdot k^{3/2} \cdot 2^{k+1}$ for every $t \in [-\sqrt{kn}, \sqrt{kn}]$,*

To prove Lemma 8, we use Lemma 4 to bound $q(t)$ on the integer points between $-\sqrt{kn}$ and $\sqrt{kn}$, and then extend the bound to the whole interval using Lemma 5.

Claim 9. *$0 \le q(t) \le \delta \cdot k^{3/2} \cdot 2^{k+1}$ for every $t \in \{-\sqrt{kn}, \dots, \sqrt{kn}\}$,*

Proof of Lemma 8. We bound $q(t)$ on $[0, \sqrt{kn}]$. A similar argument handles $[-\sqrt{kn}, 0]$. Let Q be the degree-k polynomial $Q(t) := q(\sqrt{kn} - t)$. By Claim 9, we have $|Q(j)| \le M$ on $j \in \{0, \dots, \sqrt{kn}\}$, where $M := \delta \cdot k^{3/2} \cdot 2^{k+1}$. As $3k^2 \le \sqrt{kn}$ for $k \le (n/9)^{1/3}$, by Claim 9, we have $q(t) \le 3M/2$ for every $t \in [0, \sqrt{kn}]$. □

Proof of Claim 9. Fix a point $t_0 \in [-\sqrt{kn}, \sqrt{kn}]$ and without loss of generality assume $t_0 > 0$. As $\mathbb{P}[B = t]$ is decreasing on positive integers t and $q(t) \ge 0$ on the support of B, we have

$$\begin{aligned}\mathbb{P}\Big[B = \sqrt{kn}\Big] \sum_{j=1}^{\sqrt{nk}} q(t_0 - j) &\le \sum_{j=1}^{\sqrt{nk}} \mathbb{P}\big[B = t_0 - j\big] q(t_0 - j) \\ &\le \mathbb{E}[q(B)] = \delta.\end{aligned}$$

Rearranging gives

$$\sum_{j=1}^{\sqrt{nk}} q(t_0 - j) \le \frac{\delta}{\mathbb{P}[B = \sqrt{kn}]} \le \delta \cdot \sqrt{n} \cdot 2^{k+1}.$$

Consider the polynomial $Q(t) := q(t - t_0)$ of degree k. Let $m = \sqrt{nk}$, and $L = m/k^2 = n^{1/2}k^{-3/2}$. As $k < 7\sqrt{m/L} = 7k$, by the contrapositive of Lemma 4, we have

$$\begin{aligned}|q(t_0)| = |Q(0)| &\le \frac{1}{L} \sum_{j=1}^{m} |Q(j)| \\ &= \frac{1}{L} \sum_{j=1}^{\sqrt{kn}} q(t_0 - j) \\ &\le \delta \cdot \frac{\sqrt{n} \cdot 2^{k+1}}{L} = \delta \cdot k^{3/2} \cdot 2^{k+1}.\end{aligned}$$

□

Let $s := \sqrt{kn}$. Observe that $q(t) \geq \mathbb{1}(t \geq t) = 1$. Let $\widetilde{q}_s(\theta) = q(\theta s)$. By Lemma 8, we have $\max_{\theta\in[-1,1]}|\widetilde{q}(\theta)| \leq 2\delta \cdot k^{3/2} \cdot 2^{k+1}$. By Fact 6, for $t \geq s$, we have

$$\begin{aligned} 1 \leq q(t) &= \widetilde{q}_s\Big(\frac{t}{s}\Big) \\ &\leq \Big(\frac{2t}{s}\Big)^k \cdot \max_{\theta\in[-1,1]}|\widetilde{q}_s(\theta)| \\ &\leq \Big(\frac{2t}{\sqrt{kn}}\Big)^k \cdot 2\delta \cdot k^{3/2} \cdot 2^{k+1}. \end{aligned}$$

Rearranging gives $\delta \geq \frac{1}{4k^{3/2}}(\frac{kn}{16t^2})^{k/2}$.

3 Proof of Corollary 3

Let D be the symmetrization of the k-wise uniform distribution in Theorem 2 with the t replaced with $3t$. Let D' be D conditioned on $|1^\top D| \leq n^{0.9}$. By Corollary 11, D' is $\delta := (k/n^{0.8})^{k/2}$-close to D.

Consider $D' \cdot N_{\frac{1}{2}}$. We have $|\mathbb{E}[D'^S]| \leq \delta$ for $|S| \leq k$. For $|S| \in [k+1, n]$, by Corollary 13, we have $|\mathbb{E}[D'^S]| \leq n^{-0.1k}$. For $|S| \geq n^{0.9}$, we have $|\mathbb{E}[N_{1/2}]| \leq 2^{-n^{0.9}}$.

By Theorem 2, we have $\mathbb{P}[1^\top D' \geq 3t] \geq (ckn/t^2)^{k/2} + \delta$. Note that $\mathbb{E}[1^\top(x \cdot N_\rho)] = \rho(1^\top x)$. So conditioned on $D' \geq 3t$, by a Chernoff bound, the probability $D' \cdot N_{1/2} \geq t$ is $1 - 2^{-\Omega(t)} \geq 1/2$. Hence $\mathbb{P}[1^\top(D' \cdot N_\rho) \geq t] \geq (ckn/t^2)^{k/2} + \delta \geq (ckn/t^2)^{k/2}$.

4 Proof of Theorem 1

We rely on the result from [BHLV19] that for every a, m, and $k \leq n/(8m^2)$, there is a k-wise uniform distribution D supported on $\{x \in \{-1,1\}^n : 1^\top x \equiv a \pmod m\}$. Moreover, implicit in the proof they show that the probability mass on every point s in the support of D is at least $(m/4) \cdot \mathbb{P}[1^\top U = s]$.

Let $m = \sqrt{n/(8k)}$. We can pick a such that t belongs to the support of some k-wise uniform D, from which we conclude that

$$\epsilon := \mathbb{P}[1^\top D = t] - \mathbb{P}[1^\top U = t] \geq (m/4 - 1)\,\mathbb{P}[1^\top U = t].$$

Now, we have either

$$\begin{aligned} &\mathbb{P}[1^\top D \geq t] - \mathbb{P}[1^\top U \geq t] \geq \epsilon/2 \text{ or} \\ &\mathbb{P}[1^\top D \leq t] - \mathbb{P}[1^\top U \leq t] \geq \epsilon/2, \end{aligned}$$

as otherwise, summing both inequalities give $(1 + \mathbb{P}[1^\top D = t]) - (1 + \mathbb{P}[1^\top U = t]) < \epsilon$, a contradiction. If we are in the second case, we can consider $\overline{D}$, the complement of D, which is also k-wise independent, and $\mathbb{P}[1^\top D \leq t] = \mathbb{P}[1^\top(\overline{D}) \geq t]$.

5 Alternative proofs that small-bias plus noise does not fool thresholds

We begin with some preliminaries. We need the following well-known moment bounds for k-wise uniform distributions. For a short proof see [BHLV19], Lemma 32.

Lemma 10. *Let D be a $(2k)$-wise uniform distribution on $\{-1,1\}^n$. Then $\mathbb{E}[(\sum_{i=1}^n D_i)^{2k}] \le \sqrt{2}\,(2kn/e)^k$.*

By Markov's inequality, this implies the following tail bound.

Corollary 11. *Let D be a $(2k)$-wise uniform distribution on $\{-1,1\}^n$. For every integer $t > 0$, we have*

$$\mathbb{P}\big[|1^\top D| \ge t\big] \le \sqrt{2}\left(\frac{2kn}{et^2}\right)^k.$$

The following fact says that a distribution remains close to itself after conditioning on any high probability event.

Fact 12. *Let D be any distribution on $\{-1,1\}^n$ and E be any event. Then the conditional distribution $D \mid E$ is $(1-\mathbb{P}[E])$-close to D.*

Proof. Let $\overline{E}$ be the complement of E. For every Boolean test $g\colon \{-1,1\}^n \to \{0,1\}$ we have

$$\begin{aligned}\mathbb{E}[g(D)] &= \mathbb{E}[g(D \mid E)](1-\mathbb{P}[\overline{E}]) + \mathbb{E}[g(D \mid \overline{E})]\,\mathbb{P}[\overline{E}] \\ &= \mathbb{E}[g(D \mid E)] + \big(\mathbb{E}[g(D \mid \overline{E})] - \mathbb{E}[g(D \mid E)]\big)\,\mathbb{P}[\overline{E}].\end{aligned}$$

So $|\mathbb{E}[g(D)] - \mathbb{E}[g(D \mid E)]| \le \mathbb{P}[\overline{E}]$, as $|\mathbb{E}[g(D \mid \overline{E})] - \mathbb{E}[g(D \mid E)]|$ is bounded by 1. □

The (shifted) Krawtchouk polynomials $\overline{K}$ can be defined by

$$\overline{K}(k,t) := \sum_{|S|=k} z^S,$$

where $z \in \{-1,1\}^n$ is any string such that $1^\top z = t$, and z^S is the product of the bits of z indexed by S.

For discussion and the next bound see [DILV24].

Corollary 13. *For every $1 \le \ell \le n$, we have $\big|K(\ell,t)\big| \le \binom{n}{\ell}\big(\frac{\ell}{n} + \frac{t^2}{n^2}\big)^{\frac{\ell}{2}}$.*

The next lemma is a general lemma showing that adding noise to a symmetric nearly k-wise uniform distribution causes it to be small-biased.

Lemma 14. *Let D be a symmetric distribution supported on $\{x \in \{-1,1\}^n : |1^\top x| \le t\}$ which is ϵ-close to being $(2k)$-wise uniform. Then $D \cdot N_{1/2}$ is $\big((2/n)^k + (2t/n)^{2k} + 2^{-\sqrt{n}} + \epsilon\big)$-biased.*

Proof. Let $D' = D \cdot N_{1/2}$. Note that noise does not increase bias. If $|S| \le 2k$, then $\mathbb{E}[D'^S] \le \mathbb{E}[D^S] \le \epsilon$. If $2k \le |S| \le \sqrt{n}$, then by Corollary 13 we have

$$|\mathbb{E}[D'^S]| \le |\mathbb{E}[D^S]| \le \frac{K(|S|,t)}{\binom{n}{|S|}} + \epsilon \le \left(\frac{|S|}{n} + \frac{t^2}{n^2}\right)^{|S|/2} + \epsilon \le \left(\frac{2}{n}\right)^k + \left(\frac{2t}{n}\right)^{2k} + \epsilon.$$

Finally, if $|S| \ge \sqrt{n}$, then $|\mathbb{E}[D'^S]| \le |\mathbb{E}[N_{1/2}^S]| \le 2^{-\sqrt{n}}$. □

5.1 Sums of independent random variables and normal distribution

In this section we collect several results about normal and binomial distributions that are used later. For context, we begin with the well-known Berry–Esseen theorem. Let $\mathcal{N}(0,1)$ denote the standard normal distribution, which has mean 0 and variance 1.

Lemma 15 (Theorem 11.2 in [Das08]). *Let $Y_1, \ldots, Y_n$ be n independent random variables with $\mathbb{E}[Y_i] = 0$, $\mathbf{Var}[Y_i] = \sigma_i^2$, $\mathbb{E}[|Y_i|^3] < \infty$. Let $Y := \sum_{i=1}^n Y_i$. For every $\theta \in \mathbb{R}$,*

$$\left| \mathbb{P}\left[\frac{Y}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} \geq \theta \right] - \mathbb{P}\Big[\mathcal{N}(0,1) \geq \theta \Big] \right| \leq \frac{\sum_{i=1}^n \mathbb{E}[|Y_i|^3]}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}}.$$

For fixed ρ_i's and σ_i's, the additive error given by Lemma 15 is roughly $1/\sqrt{n}$. So, for Theorem 21 to hold with $k \geq c \log n$, we would need a more refined approximation. For this reason, we will be using the following Cramér's estimate of sums of independent random variables, which gives a multiplicative rather than additive approximation in terms of the standard normal distribution.

Lemma 16 (Chapter VIII, Equation (2.41) in [Pet75]). *There exists a constant $c > 0$ such that the following holds. Let $Y_1, \ldots, Y_n$ be n independent random variables with $\mathbb{E}[Y_i] = 0$, and $\mathbb{E}[Y_i^2] = \sigma_i^2$ for each $i \in [n]$. Let $Y := \sum_{i=1}^n Y_i$. For $0 \leq \theta \leq cn^{1/6}$, there exists an $\epsilon \in [0, \frac{(\theta+1)}{c\sqrt{n}}]$ such that*

$$\mathbb{P}\left[\frac{Y}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} \geq \theta \right] = \mathbb{P}\Big[\mathcal{N}(0,1) \geq \theta \Big] \cdot \exp\left(\frac{\sum_{i=1}^n \mathbb{E}[Y_i^3]}{6\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} \cdot \theta^3 \right)(1+\epsilon).$$

To relate Lemma 16 to Lemma 15, note that when θ is small, $\exp(\frac{\sum_{i=1}^n \mathbb{E}[Y_i^3]}{6(\sum_{i=1}^n \sigma_i^2)^{3/2}} \cdot \theta^3)$ is roughly $1 + \frac{\sum_{i=1}^n \mathbb{E}[Y_i^3]}{(\sum_{i=1}^n \sigma_i^2)^{3/2}}$. We will use following approximation on the tails of the standard normal distribution.

Lemma 17 (Lemma 22.2 in [Kle20]). *For any $\theta > 0$,*

$$\frac{1}{\theta + \frac{1}{\theta}} \leq \mathbb{P}\big[\mathcal{N}(0,1) \geq \theta\big] \cdot \frac{\sqrt{2\pi}}{e^{-\theta^2/2}} \leq \frac{1}{\theta}.$$

5.2 Three more ways to get bounded-uniform distributions with contrained support

The first is given immediately from this result.

Lemma 18 ([BHLV19]). *For any integer k, there is a symmetric distribution D supported on $\{x \in \{-1,1\}^n : |1^\top x| \leq 10\sqrt{kn}\}$ which is $(2k)$-wise uniform.*

For the second, rather then relying on the somewhat technical result in [BHLV19], we use Carathéodory's theorem from convex geometry, stated next, which has a simple proof. See e.g. Theorem 2.3 in Chapter 1 in [Bar02]:

Lemma 19. *Let $S \subset \mathbb{R}^d$ be a set. Every point in the convex hull of S can be represented as a convex combination of $d+1$ points from S.*

Next is our application of this lemma in this context.

Lemma 20. *For any integer k and t, there is a symmetric distribution D supported on $\{x \in \{-1,1\}^n : 1^\top x \in \{w_1, \ldots, w_{2k+1}\} \wedge |w_1|, \ldots, |w_{2k+1}| \leq t\}$ which is $(ckn/t^2)^{k/2}$-close to a $(2k)$-wise uniform distribution.*

Proof. Let $Z := \{-n, -n+1, \ldots, n\}$. For each $i \in Z$ define $v_i := (i, i^2, \ldots, i^k) \in \mathbb{R}^k$ and define $b \in \mathbb{R}^k$ as $b := \sum_{i \in Z} \mathbb{P}[B = i] v_i = (\mathbb{E}[B], \mathbb{E}[B^2], \ldots, \mathbb{E}[B^k])$, which is manifestly in the convex hull of the v_i. By Lemma 19 b is a convex combination of $k+1$ points v_i. Let $W \subset Z$ be the corresponding set of $k+1$ indices i. So then we can write $b = \sum_{i \in W} p_i v_i$ where the p_i are ≥ 0 and sum to 1.

Let D_W be the distribution over W where $\mathbb{P}[D_W = i] = p_i$. We have $(\mathbb{E}[D_W], \mathbb{E}[D_W^2], \ldots, \mathbb{E}[D_W^k]) = b$ and so the first k moments of D_W match those of B. Let D_k be the symmetric distribution over $\{-1,1\}^n$ such that $\mathbb{P}[1^\top(D_k) = i] = \mathbb{P}[D_W = i]$. Note that D_k is k-wise uniform. We conclude by taking D_k conditioned on the event $1^\top(D_k) < t$, which is $(ckn/t^2)^{k/2}$-close to D_k by Corollary 11 and Fact 12. □

Finally, we mention that a third way is to use the standard construction of bounded-uniform distributions via BCH codes, and then use results from algebraic geometry to bound the Hamming weight, specifically Theorem 18 in [MS77].

5.3 BHLV

Here we prove Theorem 21, which immediately follows from Lemma 22.

Theorem 21. *For any $\rho \in (0,1]$ and $k \leq c\rho^2 n^{1/3}$, there is a n^{-ck}-biased distribution D on $\{-1,1\}^n$ and $\theta = c\sqrt{nk}/\rho$ such that*

$$\mathbb{P}[B \geq \theta] \geq \mathbb{P}[1^\top(D \cdot N_\rho) \geq \theta] + 2^{-ck/\rho^2}.$$

Lemma 22. *Let k and $\rho \in (0,1]$ such that $k \leq c\rho^2 n^{1/3}$. Let $\theta = c\sqrt{nk}/\rho$. For every $x \in \{-1,1\}^n$ with $|1^\top x| \leq 10\sqrt{nk}$ we have*

$$\mathbb{P}\big[B \geq \theta\big] \geq \mathbb{P}\big[1^\top(x \cdot N_\rho) \geq \theta\big] + 2^{-ck/\rho^2}.$$

Proof of Theorem 21 assuming Lemma 22. Let $D := D' \cdot N_{1/2}$, where D' is the $(2k)$-wise uniform distribution from Lemma 18 supported on $\{x \in \{-1,1\}^n : |1^\top x| \leq 10\sqrt{kn}\}$. By Lemma 14, D is $(ck/n)^k$-biased. The theorem now follows by Lemma 22, as $D \cdot N_\rho = D' \cdot N_{1/2} \cdot N_\rho = D' \cdot N_{\rho/2}$. □

Proof of Lemma 22. Fix any $x \in \{-1,1\}^n$. Note that for any $i \in [n]$,

$$\mathbb{E}[(x \cdot N_\rho)_i] = \rho x_i, \quad \mathbb{E}[(x \cdot N_\rho)_i^2] = 1 \quad \text{and} \quad \mathbb{E}[(x \cdot N_\rho)_i^3] = -\rho x_i$$

Let us consider the mean zero variables $Y_i := (x \cdot N_\rho)_i - \rho x_i$. We have

$$\mathbb{E}[Y_i] = 0, \quad \sigma_\rho^2 := \mathbb{E}[Y_i^2] = 1 - \rho^2, \quad \text{and} \quad \mathbb{E}[Y_i^3] = 2\rho(\rho^2 - 1)x_i = -2\rho \cdot \sigma_\rho^2 \cdot x_i.$$

Now fix any $x : |1^\top x| \le 10\sqrt{nk} := t$. Define $\theta = \beta\sqrt{nk}/\rho$ for some parameter β. Since $\sigma_\rho^{-1} = \frac{1}{\sqrt{1-\rho^2}} \ge 1 + \frac{\rho^2}{2}$,

$$\frac{\theta - \rho 1^\top x}{\sigma_\rho \sqrt{n}} \ge \left(1 + \frac{\rho^2}{2}\right)\frac{\theta - \rho t}{\sqrt{n}} \ge \beta\sqrt{k}\left(\frac{1}{\rho} + \frac{\rho}{4}\right)$$

where the last inequality holds for $\beta > 60$. Thus we have

$$\begin{aligned}
\mathbb{P}\Big[1^\top(x \cdot N_\rho) \ge \theta\Big] &= \mathbb{P}\left[\frac{\sum_{i=1}^n Y_i}{\sigma_\rho\sqrt{n}} \ge \frac{\theta - \rho 1^\top x}{\sigma_\rho\sqrt{n}}\right] \\
&\le \mathbb{P}\left[\frac{\sum_{i=1}^n Y_i}{\sigma_\rho\sqrt{n}} \ge \beta\sqrt{k}\left(\frac{1}{\rho} + \frac{\rho}{4}\right)\right] \\
&\le 4\,\mathbb{P}\left[\mathcal{N}(0,1) \ge \beta\sqrt{k}\left(\frac{1}{\rho} + \frac{\rho}{4}\right)\right].
\end{aligned}$$

The last $\le$ follows from Lemma 16 and bounds on the error terms which we justify at the end. And again by Lemma 16,

$$\mathbb{P}\big[B \ge \theta\big] = \mathbb{P}\left[\frac{B}{\sqrt{n}} \ge \frac{\theta}{\sqrt{n}}\right] \ge \mathbb{P}\left[\mathcal{N}(0,1) \ge \frac{\beta\sqrt{k}}{\rho}\right].$$

Next, using both sides of Lemma 17,

$$\frac{\mathbb{P}\left[\mathcal{N}(0,1) \ge \frac{\beta\sqrt{k}}{\rho}\right]}{4\,\mathbb{P}\left[\mathcal{N}(0,1) \ge \beta\sqrt{k}\left(\frac{1}{\rho} + \frac{\rho}{4}\right)\right]} \ge ce^{-\beta^2 k/8}.$$

Setting β to a large enough constant concludes the proof.

It remains to bound the error term from Lemma 16. We claim

$$\exp\left(\frac{-2\rho\sigma_\rho^2 1^\top x}{6(\sigma_\rho\sqrt{n})^3} \cdot \left(\frac{c\beta\sqrt{k}}{\rho}\right)^3\right) \cdot \left(1 + \frac{c\beta\sqrt{k}}{\rho\sqrt{n}}\right) \le 4.$$

Since $k \le c\rho^2 n^{1/3}$, the right term is ≤ 2 and the first term is $\le 1 + \frac{c\beta^3 k^2}{\rho^2\sigma_\rho n} \le 2$ since $1^\top x \ge -t$ and $\exp(x) \le 1 + 2x$ for $0 < x < 1$. □

5.4 Caratheodory and mixture of gaussians

Next we give an alternate version of Theorem 21, stated below.

Theorem 23. *For any $\rho \in (0,1]$ and $k \le cn^{1/2}$ there is a n^{-ck}-biased distribution D on $\{-1,1\}^n$ and some θ such that*

$$|\,\mathbb{P}[B \ge \theta] - \mathbb{P}[1^\top(D \cdot N_\rho) \ge \theta]| \ge 2^{-ck/\rho} - \frac{c}{\sqrt{n}}.$$

The main technical result we need, which may be of independent interest, is stated next. We defer its proof to the next section.

Lemma 24. *Let M be a mixture of k Gaussian distributions each with variance $\sigma^2 = 1 - \rho^2$. Then there exists an interval I such that*

$$|\,\mathbb{P}[\mathcal{N}(0,1) \in I] - \mathbb{P}[M \in I]| \geq 2^{-ck/\rho}.$$

In particular, up to a factor 2 the same bound applies to some threshold.

Proof of Theorem 23 assuming Lemma 24. Let D' denote the distribution from Lemma 20 with $t = k^{1/4}n^{3/4}$, and define $D = D' \cdot N_{1/2}$ By Lemma 14, D is ϵ-biased where

$$\epsilon = \left(\frac{ct}{n}\right)^{2k} + \left(\frac{ckn}{t^2}\right)^{k} = \left(\frac{ck}{n}\right)^{k/2}.$$

Note $D \cdot N_\rho = D' \cdot N_{\rho/2}$. Now we claim that by the Berry-Essen theorem (Lemma 15), $D' \cdot N_{\rho/2}$ is $(cn^{-1/2})$ close to a mixture of $2k+1$ Gaussian distributions, each with variance $1 - \rho^2/4$. To see this, note for any $i \in [n]$,

$$\mathbb{E}[(D' \cdot N_{\rho/2})_i] = (\rho/2)D'_i, \quad \mathbb{E}[(D' \cdot N_{\rho/2})_i^2] = 1 \quad \text{and} \quad \mathbb{E}[(D' \cdot N_{\rho/2})_i^3] = -(\rho/2)D'_i.$$

Next we consider the mean zero variables $Y_i := (D' \cdot N_{\rho/2})_i - (\rho/2)D'_i$. We have

$$\mathbb{E}[Y_i] = 0, \quad \mathbb{E}[Y_i^2] = 1 - \rho^2/4, \quad \text{and} \quad \mathbb{E}[|Y_i|^3] \leq c.$$

Recall D' is supported on $\leq 2k+1$ distinct weights. Condition on D' being fixed to one of these weights. Then for any θ, by Lemma 15 we have

$$\begin{aligned}\mathbb{P}\Big[1^\top(D' \cdot N_{\rho/2}) \geq \theta\Big] &= \mathbb{P}\left[\frac{\sum_{i=1}^n Y_i}{\sqrt{n}} \geq \frac{\theta - (\rho/2)1^\top D'}{\sqrt{n}}\right] \\ &\leq \mathbb{P}\Big[\mathcal{N}(\mu, 1 - \rho^2/4) \geq \theta\Big] + \frac{c}{\sqrt{n}}.\end{aligned}$$

Note μ only depends on the weight of D'. Thus $D' \cdot N_{\rho/2}$ is $(c/\sqrt{n})$-close to a mixture of $2k+1$ Gaussians each with variance $1 - \rho^2/4$.

On the other hand, again by Lemma 15,

$$\mathbb{P}\big[B \geq \theta\sqrt{n}\big] = \mathbb{P}\left[\frac{B}{\sqrt{n}} \geq \theta\right] \geq \mathbb{P}\Big[\mathcal{N}(0,1) \geq \theta\Big] - \frac{c}{\sqrt{n}}.$$

Combining the above with Lemma 24, there exists some θ such that

$$\begin{aligned}\left|\mathbb{P}\big[B \geq \theta\sqrt{n}\big] - \mathbb{P}\Big[1^\top(D \cdot N_\rho) \geq \theta\sqrt{n}\Big]\right| &\geq \left|\mathbb{P}\Big[\mathcal{N}(0,1) \geq \theta\Big] - \mathbb{P}\Big[\mathcal{N}(\mu, 1 - \rho^2/4) \geq \theta\Big]\right| - \frac{c}{\sqrt{n}} \\ &\geq 2^{-ck/\rho} - \frac{c}{\sqrt{n}}.\end{aligned}$$

□

6 Proof of Lemma 24

We wish to show that a linear sum of k exponential functions with variance < 1 can not approximate the standard normal well. We can factor the two expressions so the mixture becomes a linear sum of k exponential functions, which can be written as $\sum_{i\in[k]} a_i e^{b_i}$, while the standard normal can be written as $e^{\alpha x^2}$. This factoring crucially uses the fact the variances in the mixture are identical.

We then argue the distance must be large for some point. To prove this, we show the entries in the inverse of a Vandermonde like matrix which corresponds to the $e^{\alpha x^2}$ are not too large. This step is the bulk of the proof. On the other hand, the Vandermonde matrix corresponding to the sum of exponentials is singular. After some matrix norm manipulations this allows us to achieve the desired result.

Lemma 25. *Suppose that $f(x)$ is the PDF of a Gaussian distribution with variance 1, and $g(x)$ is the PDF of a mixture of k Gaussian distributions with variance $1-\rho^2 = \sigma^2 < 1$. Then*

$$\|f - g\|_\infty \geq e^{-ck/\rho}.$$

We let $\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ denote the probability density function of $\mathcal{N}(0,1)$.

Proof of Lemma 24 assuming Lemma 25. Let $g(x)$ denote the PDF of M and set $h(x) = \phi(x) - g(x)$. By Lemma 25, there exists some $a \in \mathbb{R}$ with

$$|h(a)| \geq e^{-ck/\rho}.$$

We have $|\phi'(x)| \leq (2\pi e)^{-1/2}$ and $|g'(x)| \leq (2\pi e)^{-1/2}/\sigma^2$ for all x, so $|h'(x)| \leq (2\pi e)^{-1/2}(1 + \frac{1}{\sigma^2}) < \frac{1}{2\sigma^2}$. We claim there exists an interval $I \subseteq \mathbb{R}$ with

$$|\,\mathbb{P}[\mathcal{N}(0,1) \in I] - P[M \in I]| \geq 2\sigma^2 \left(e^{-ck/\rho}\right)^2 = e^{-ck/\rho}.$$

To see this, assume that $h(0) \geq e^{-ck/\rho}$. Then $h(x) \geq e^{-ck/\rho} - \frac{1}{2\sigma^2}|x|$ for any x. So we set the interval $I = [-2\sigma^2 e^{-ck/\rho}, 2\sigma^2 e^{-ck/\rho}]$, and then $\int_I h(x)dx \geq 2\sigma^2(e^{-ck/\rho})^2$. Finally, note the assumption that $a = 0$ can be made without loss of generality. □

6.1 Proof of Lemma 25

The main technical result we need is the following.

Lemma 26. *Let $\alpha, D > 0$ be fixed. Let*

$$\Delta(k) := \inf_g \|e^{\alpha x^2} - g(x)\|_\infty$$

where the infimum is over g that are a linear combination of k exponential functions, and the norm $\|\cdot\|_\infty$ is the supremum over the interval $[-D\sqrt{k}, D\sqrt{k}]$. Then we have

$$\Delta(k) \geq \exp\left(\frac{-ck}{D^2\alpha}\right).$$

Proof of Lemma 25 assuming Lemma 26. Without loss of generality we may assume that $f(x)$ is the PDF of the standard normal distribution with mean 0 and variance 1. Define $\overline{f}(x) = e^{x^2/2\sigma^2} f(x) = e^{\alpha x^2}$ with $\alpha = \frac{1}{2\sigma^2} - \frac{1}{2}$, and $\overline{g}(x) = e^{x^2/2\sigma^2} g(x)$. Now $g(x)$ is a linear combination of k exponential functions. If we choose some $D > 0$ then Lemma 26 gives us

$$\Delta(k) \geq \exp\Big(\frac{-ck}{D^2\alpha}\Big),$$

where $\Delta(k)$ is the supremum of $|\overline{f} - \overline{g}|$ over the interval $[-D\sqrt{k}, D\sqrt{k}]$. It follows that

$$\|f - g\|_\infty = \|e^{-x^2/2\sigma^2}(\overline{f} - \overline{g})\|_\infty \geq \exp\Big(\frac{-D^2 k}{2\sigma^2}\Big)\Delta(k) = \exp\Big(-k\Big(\frac{D^2}{2\sigma^2} + \frac{c}{D^2\alpha}\Big)\Big),$$

Then if we set $D^2 = c\sqrt{\sigma^2/\alpha}$ we have

$$\|f - g\|_\infty \geq \exp\Big(\frac{-ck}{\sqrt{\sigma^2\alpha}}\Big) = \exp\Big(\frac{-ck}{\sqrt{1-\sigma^2}}\Big). \qquad \square$$

6.2 Proof of Lemma 26

Definition 27. For a function $f : \mathbb{Z} \to \mathbb{R}$ define the $(k+1) \times (k+1)$ matrix $M_k(f)$ by $M_k(f)_{i,j} = f(i + j - k - 2)$. For example,

$$M_3(f) = \begin{pmatrix} f(-3) & f(-2) & f(-1) & f(0) \\ f(-2) & f(-1) & f(0) & f(1) \\ f(-1) & f(0) & f(1) & f(2) \\ f(0) & f(1) & f(2) & f(3) \end{pmatrix}$$

Fact 28. *If $f(x) = \sum_{i=1}^k a_i e^{b_i x}$, then $\det M_k(f) = 0$.*

Proof. We can write $f(x) = \sum_{i=1}^k a_i \lambda_i^x$ where $\lambda_i = e^{\beta_i}$. Consider the polynomial

$$p(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_k) = x^k + c_{k-1}x^{k-1} + \cdots + c_1 x + c_0.$$

Because $\lambda_1, \lambda_2, \ldots, \lambda_k$ are roots of $p(x)$, we have

$$0 = \sum_{i=1}^k a_i \lambda_i^x p(\lambda_i) = \sum_{i=1}^k a_i(\lambda_i^{x+k} + c_{k-1}\lambda_i^{x+k-1} + \cdots + c_0\lambda_i^n) =$$
$$= f(x+k) + c_{k-1}f(x+k-1) + \cdots + c_1 f(x+1) + c_0 f(x).$$

For the vector $c = [c_0\ c_1\ \cdots\ c_{k-1}\ 1]$ we get $cM_k(f) = 0$, so $\det M_k(f) = 0$. We claim the sequence $\ldots, f(-1), f(0), f(1), f(2), \ldots$ satisfies a linear recurrence of order k, which implies the columns of $M_k(f)$ are linearly dependent.

To prove the claim, we show the existence of $c_1, \ldots, c_k$ such that

$$f(x) = c_1 f(x-1) + \cdots + c_k f(x-k).$$

Solving for $c_1, \ldots, c_k$, we get

$$1 = \frac{c_1}{e^{b_1}} + \cdots + \frac{c_k}{e^{kb_1}}, \ldots, 1 = \frac{c_1}{e^{b_k}} + \cdots + \frac{c_k}{e^{kb_k}}.$$

There exists a solution to this system. $\square$

Fact 29. *If $|x| < 1$ then*

$$\prod_{i=1}^{\infty}(1-x^i) \geq \exp\left(\frac{-c}{1-x}\right).$$

Proof. Using absolute convergence, and the inequality $(1-x^j) \geq jx^j(1-x)$, we get

$$\sum_{i=1}^{\infty}\log(1-x^i) = -\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{x^{ij}}{j} = -\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\frac{x^{ij}}{j} = -\sum_{j=1}^{\infty}\frac{x^j}{j(1-x^j)} \geq -\sum_{j=1}^{\infty}\frac{1}{j^2(1-x)} = -\frac{\pi^2}{6(1-x)}.$$

Then apply the exponential function to both sides. □

Lemma 30. *Suppose that $q > 1$ and k is a positive integer. Let $A := M_k(q^{x^2})^{-1}$. Then*

$$|A_{i,j}| \leq \frac{\binom{k}{i-1}\binom{k}{j-1}}{\prod_{i=1}^{k}(1-q^{-2i})}.$$

Proof of Lemma 26. Let $f(x) = e^{\alpha x^2}$ and $g(x) = \sum_{i=1} a_i e^{b_i x}$. Define $\widetilde{f}(x) = f(Dx/\sqrt{k}) = q^{x^2}$, where $q = e^{D^2\alpha/k}$ and $\widetilde{g}(x) = g(Dx/\sqrt{k})$. By Fact 28, $M_k(\widetilde{g})$ is singular. Let A be the inverse of $M_k(\widetilde{f})$. The matrix

$$AM_k(\widetilde{g}) = AM_k(\widetilde{f}) - AM_k(\widetilde{f}-\widetilde{g}) = I - AM_k(\widetilde{f}-\widetilde{g})$$

is singular. It follows that

$$\|A\|_\sigma \|M_k(\widetilde{f}-\widetilde{g})\|_\sigma \geq \|AM_k(\widetilde{f}-\widetilde{g})\|_\sigma \geq 1$$

where $\|A\|_\sigma$ is the spectral norm of A. The matrix A is positive definite symmetric and the sum of the singular values is the sum of the eigenvalues which is equal to the trace of A. By Lemma 30 we get

$$\|A\|_\sigma \leq \mathrm{trace}(A) \leq \frac{\sum_{i=0}^{k}\binom{k}{i}^2}{\prod_{i=1}^{k}(1-q^{-2i})} = \frac{4^k}{\prod_{i=1}^{k}(1-q^{-2i})}.$$

On the other hand,

$$\|M_k(\widetilde{f}-\widetilde{g})\|_\infty \leq (k+1)\|f-g\|_\infty.$$

Combining everything and using Fact 29, we get

$$\|f-g\|_\infty \geq \frac{\prod_{i=1}^{k}(1-q^{-2i})}{4^k(k+1)} \geq \exp\left(\frac{-c}{1-q^{-2}} - \log(4)k - \log(k+1)\right) = \exp\left(\frac{-ck}{D^2\alpha}\right). \quad \square$$

6.3 Proof of Lemma 30

First we define Vandermonde matrices.

Definition 31.

$$\mathrm{Vand}(1, x_1, x_2, \ldots, x_k) = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & x_1 & x_2 & \cdots & x_k \\ 1 & x_1^2 & x_2^2 & \cdots & x_k^2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_1^k & x_2^k & \cdots & x_k^k \end{pmatrix}.$$

Proof. We can transform $M_k(q^{x^2})$ into $\mathsf{Vand}(1, q^2, q^4, \ldots, q^{2k})$ by multiplying the rows and columns of $M_k(q^{x^2})$ with powers of q. Thus by Proposition 32, stated at the end, $(-1)^{i+j} A_{i,j} \prod_{b=1}^{k} (q^{2b} - 1)$ is a sum of $\binom{k}{i-1}\binom{k}{j-1}$ powers of q. This implies $(-1)^{i+j} A_{i,j} \prod_{b=1}^{k} (1 - q^{-2b})$ is also a sum of $\binom{k}{i-1}\binom{k}{j-1}$ powers of q.

Next we claim no positive powers of q appear in the aforementioned sum. We define

$$B_k(q) := \begin{pmatrix} q^{(-k)^2/2} & & & \\ & q^{(2-k)^2/2} & & \\ & & \ddots & \\ & & & q^{k^2/2} \end{pmatrix}$$

so that we can write

$$M_k(q^{x^2}) = B_k(q) C_k(q) B_k(q)$$

where $C_k(q)$ is a matrix with 1 on the diagonal and negative powers of q outside of the diagonal. In particular, $C_k(q)$ converges to the identity matrix as $q \to \infty$. So

$$A = A(q) = M_k(q^{x^2})^{-1} = B_k(q)^{-1} C_k(q)^{-1} B_k(q)^{-1}$$

converges as $q \to \infty$ because both $B_k(q)^{-1}$ and $C_k(q)^{-1}$ converge. This shows that $A_{i,j} \prod_{b=1}^{k} (1 - q^{-2b})$ cannot have positive powers of q in its corresponding sum. Since

$$|A_{i,j}| \prod_{b=1}^{k} (1 - q^{-2b})$$

is a sum of $\binom{k}{i-1}\binom{k}{j-1}$ non-positive powers of q and $q > 1$ we get

$$|A_{i,j}| \prod_{b=1}^{k} (1 - q^{-2b}) \leq \binom{k}{i-1}\binom{k}{j-1}.$$ □

Proposition 32. *Let* $V = \mathrm{Vand}(1, q, q^2, \ldots, q^k)$. *Then*

$$(-1)^{i+j} (V^{-1})_{i,j} \prod_{b=1}^{k} (q^b - 1)$$

is a sum of $\binom{k}{i-1}\binom{k}{j-1}$ *powers of* q.

Proof. Note that $\det(V) = \prod_{0\le a<b\le k}(q^b - q^a)$. Let $\widetilde{V}_i$ be the matrix V with the i-th column removed, and $\widetilde{V}_{j,i}$ be the matrix V with the j-th row and i-th column removed. By the formula of V^{-1} from Cramer's rule we get

$$(V^{-1})_{i,j} = \frac{(-1)^{i+j}\det(\widetilde{V}_{j,i})}{\det(V)}.$$

Note that $\widetilde{V}_{k+1,i} = \mathrm{Vand}(1, q, \cdots, q^{i-2}, q^i, \ldots, q^{k+1})$, so

$$\det(\widetilde{V}_{k+1,i}) = \prod_{\substack{0\le a<b\le k\\ a,b\neq i-1}} (q^b - q^a).$$

So we have

$$\frac{\prod_{b=1}^{k}(q^b-1)\det(\widetilde{V}_{k+1,i})}{\det(V)} = \frac{\prod_{b=1}^{k}(q^b-1)}{\prod_{j=i}^{k}(q^j-q^{i-1})\prod_{j=0}^{i-2}(q^{i-1}-q^j)}$$

which is up to a power of q factor equal to

$$\frac{\prod_{j=1}^{k}(q^j-1)}{\prod_{j=1}^{k-i+1}(q^j-1)\prod_{j=1}^{i-1}(q^j-1)} = \binom{k}{i-1}_q,$$

where the right-hand side is a Gaussian q-binomial coefficient which is a sum of $\binom{k}{i-1}$ powers of q. To see this, consider the generating function $\prod_{j=0}^{k-1}(1+q^jt) = \sum_{j=0}^{k} q^{j(j-1)/2}\binom{k}{j}_q t^j$. This implies that $\binom{k}{i-1}_q$ is a sum of $\binom{k}{i-1}$ powers of q.

The quotient

$$\frac{\det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_k \\ \vdots & \vdots & & \vdots \\ x_1^{j-2} & x_2^{j-2} & \cdots & x_k^{j-2} \\ x_1^{j} & x_2^{j} & \cdots & x_k^{j} \\ \vdots & \vdots & & \vdots \\ x_1^{k} & x_2^{k} & \cdots & x_k^{k} \end{pmatrix}}{\det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_k \\ x_1^2 & x_2^2 & \cdots & x_k^2 \\ \vdots & \vdots & & \vdots \\ x_1^{k-1} & x_2^{k-1} & \cdots & x_{k-1}^{k} \end{pmatrix}}$$

is equal to the elementary symmetric polynomial $e_{k+1-j}(x_1, x_2, \ldots, x_k)$ which is the sum of all $\binom{k}{k+1-j} = \binom{k}{j-1}$ squarefree monomials in $x_1, x_2, \ldots, x_k$. [1]

[1]This can be seen as follows. We start by repeating a proof for the determinant of the regular Vandermonde. Note $(x_j - x_i)$ is a factor of the determinant, since if we replace x_j with x_i the determinant becomes 0. Thus

If we apply this to the case where $x_1, x_2, \ldots, x_k$ are $1, q, \ldots, q^{i-2}, q^i, \ldots, q^k$ then we see that

$$\frac{\det(\widetilde{V}_{j,i})}{\det(\widetilde{V}_{k+1,i})} = e_{k+1-j}(1, q, \ldots, q^{i-2}, q^i, \ldots, q^k)$$

is a sum of $\binom{k}{j-1}$ powers of q. We conclude that

$$(-1)^{i+j}(V^{-1})_{i,j}\prod_{b=1}^{k}(q^b-1) = \frac{\det(\widetilde{V}_{j,i})}{\det(\widetilde{V}_{k+1,i})} \cdot \frac{\prod_{b=1}^{k}(q^b-1)\det(\widetilde{V}_{k+1,i})}{\det(V)}.$$

is a sum of $\binom{k}{i-1}\binom{k}{j-1}$ powers of q.

□

References

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized algorithm for the maximal independent set problem. *Journal of Algorithms*, 7:567–583, 1986. 1

[Bar02] Alexander Barvinok. *A course in convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002. 5.2

[BGGP12] Itai Benjamini, Ori Gurel-Gurevich, and Ron Peled. On k-wise independent distributions and boolean functions, 2012. ??, 1.1

[BHLV19] Ravi Boppana, Johan Håstad, Chin Ho Lee, and Emanuele Viola. Bounded independence versus symmetric tests. *ACM Trans. Comput. Theory*, 11(4):Art. 21, 27, 2019. 7, 4, 5, 18, 5.2

[BS15] Mark Bun and Thomas Steinke. Weighted polynomial approximations: limits for learning and pseudorandomness. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 40 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 625–644. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2015. 1.1, 2, 2, 5

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions (preliminary version). In *26th Symposium on Foundations of Computer Science*, pages 396–407, Portland, Oregon, 21–23 October 1985. IEEE. 1

[CW79] J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *J. of Computer and System Sciences*, 18(2):143–154, 1979. 1

the determinant can be expressed as $p\prod_{1\le b<a\le k}(x_b - x_a)$ for some polynomial p. Note the determinant is homogenous of degree $(k)(k+1)/2$ thus p must be 1. Now, we replace degree $j-1$ with degree k. p must be symmetric (since swapping x_b, x_a only affects the sign, and this is reflected in $\prod_{1\le b<a\le k}(x_b - x_a)$. Moreover, p is squarefree since otherwise the degree of a single variable in each monomial becomes $> k$, which is a contradiction.

[Das08] Anirban DasGupta. *Asymptotic theory of statistics and probability*. Springer Texts in Statistics. Springer, New York, 2008. 15

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM J. on Computing*, 39(8):3441–3462, 2010. **??**

[DILV24] Harm Derksen, Peter Ivanov, Chin Ho Lee, and Emanuele Viola. Pseudorandomness, symmetry, smoothing: I. In *Conf. on Computational Complexity (CCC)*, 2024. 1, 1.2, 1.3, 1.3, 1.3, 5

[DKN10] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010*, pages 11–20. IEEE Computer Soc., Los Alamitos, CA, 2010. **??**

[Erd16] Tamás Erdélyi. Coppersmith-Rivlin type inequalities and the order of vanishing of polynomials at 1. *Acta Arith.*, 172(3):271–284, 2016. 4

[HH23] Pooya Hatami and William Hoza. Theory of unconditional pseudorandom generators. *Electron. Colloquium Comput. Complex.*, TR23-019, 2023. 1

[Kle20] Achim Klenke. *Probability theory—a comprehensive course*. Universitext. Springer, Cham, [2020] ©2020. Third edition [of 2372119]. 17

[LV17] Chin Ho Lee and Emanuele Viola. Some limitations of the sum of small-bias distributions. *Theory of Computing*, 13, 2017. 1.2

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. II*. North-Holland Mathematical Library, Vol. 16. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977. 5.2

[NN90] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. In *22nd ACM Symp. on the Theory of Computing (STOC)*, pages 213–223. ACM, 1990. 1

[Pet75] V. V. Petrov. *Sums of independent random variables*. Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], Band 82. Springer-Verlag, New York-Heidelberg, 1975. Translated from the Russian by A. A. Brown. 16

[RR47] C. Radhakrishna Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *Suppl. J. Roy. Statist. Soc.*, 9:128–139, 1947. 1